PATENT
ATTORNEY DOCKET NO.: INGA,004/CIP

APPLICATION FOR UNITED STATES LETTERS PATENT

Entitled

IN VIVO PRODUCTION OF ssDNA USING REVERSE TRANSCRIPTASE WITH PREDEFINED REACTION TERMINATION VIA STEM-LOOP FORMATION

Applicant: Charles A. Conrad
Over Land Park, Kansas

09397603 .091699

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Number: EE874591908US
Date of Deposit: Sept. 16, 1999

I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR §1.10 on the date indicated above and is addressed to Box Patent Application, The Honorable Commissioner of Patent and Trademarks, Washington, D.C. 20231

Mark R. Wisner, Registration No. 30,603

September 16, 1999
Date of Signature

IN VIVO PRODUCTION OF ssDNA USING REVERSE TRANSCRIPTASE WITH PREDEFINED REACTION TERMINATION VIA STEM-LOOP FORMATION

BACKGROUND OF THE INVENTION

This application is a continuation-in-part of my co-pending application Serial No. 09/169,793, entitled PRODUCTION OF ssDNA *IN VIVO*, filed October 9, 1998. Serial No. 09/169,793 is itself a continuation-in-part of application Serial No. 08/877,251, entitled STEM-LOOP CLONING VECTOR AND METHOD, filed June 17, 1997. Serial No. 08/877,251 is a continuation application of application Serial No. 08/236,504, having the same title, filed April 29, 1994.

The present invention relates to production of single-stranded DNA (ssDNA) from a plasmid-based system within eukaryotic cells. More specifically, the present invention relates to methods and compositions for expression of any ssDNA within eukaryotic host cells with the unique ability to remove contiguous plasmid vector sequences by stem-loop secondary structure with subsequent permature termination of a reverse transcription reaction.

So far as is known, there is no method for producing single-stranded deoxyribonucleic acid (ssDNA) species in eukaryotic cells which do not contain intervening and/or flanking vector sequences. The scientific and patent literature does include the disclosure of cDNA-producing vectors (*see* A. Ohshima, *et al.*, 89 Proc. Natl. Acad. Sci. USA 1016-1020 (1992); S. Inouye, *et al.*, 3 Current Opin. Genet. Develop. 713-718 (1993); O. Mirochnitchenko, *et al.*, 269 J. Biol. Chem. 2380-2383 (1994); J.-R. Mao, *et al.*, 270 J. Biol. Chem. 19684-19687 (1995); and U.S. Patent No. 5,436,141), but that system does not appear to have demonstrated the ability to produce ssDNA in eukaryotic cells without intervening vector sequences. The ssDNA produced by the method described in these references includes intervening nucleotide sequences which can interfere with the intended function of the ssDNA product.

There are also a number of viral and transposable elements that have been discovered which contain ssDNA intermediates within their life cycles in prokaryotic systems and yeast (*see* A.M. Weiner, *et al.*, 55 Ann. Rev. Biochem. 631-661 (1986) and

H. Varmus, *et al.*, *in* Mobile DNA, M.M. Howe and D.E. Berg (Eds.), American Society for Microbiology: Washington, D.C., pp. 53-108 (1989)). Many of these genetic elements can be adapted to produce single-stranded nucleic acids within eukaryotic systems but would necessarily contain genetic information (nucleotide bases) which could interfere with the desired function of the *in vivo* produced single-stranded oligonucleotides or are difficult to manipulate beyond their normal biological life cycles to carry desired nucleotide sequences into the cell. The RNA expression vectors which have been described in the literature produce antisense RNA in cells, but the RNA:RNA hybrids which are formed are not as stable as DNA:RNA hybrids, the synthesized RNA has a shorter half-life than DNA, and RNA-RNA hybrids do not stimulate RNase H activity, which may be critical for *in vivo* antisense efficacy, as does the DNA:RNA hybrids (H. Donis-Keller, 7 Nucleic Acids Res. 179-192 (1979)).

Artificially synthesized DNA analog oligomers used for antisense therapies must be administered intravenously, which involves problems in cell uptake and distribution (P.A. Cossum, *et al.*, 267 J. Pharmacol. Expl. Ther. 1181-1190 (1993); H. Sands, *et al.*, 47 Mol. Pharmacol. 636-646 (1995)) as well as toxicity problems due to the high blood concentrations required to be effective (S.P. Henry, *et al.*, 116 Toxicology 77-88 (1997)). By far the most used DNA analogs in antisense therapies are phosphorothioates and methylphosphonates. However, phosphorothioate oligonucleotides tend to bind serum and intracellular proteins nonspecifically (S.T. Crooke, *et al.*, 227 J. Pharmacol. Exp. Ther. 923-937 (1996); W.Y. Gao, *et al.*, 41 Mol. Pharmacol. 223-229 (1992)), and at higher concentrations, inhibit RNase H activity (S.T. Crooke, *et al.*, 312 Biochem. J. 599-608 (1995)). Phosphorothioate oligonucleotides have a lower Tm (an average of 0.50C per base-pair) for RNA than does DNA (S.T. Crooke, *et al.*, Antisense Research and Application, CRC Press: Boca Raton (1993)), which requires that phosphorothioate oligonucleotides be typically longer than phoshodiester DNA oligonucleotides for effective binding, which can cause a loss of hybridization specificity (J.-J. Toulmé, *et al.*, *in* C. Lichtenstein and W. Nellen (Eds.), Antisense Technology: A Practical Approach IRL Press: New York, pp. 39-74 (1997)). Further, although their pharmacokinetic properties appear to be more favorable than unmodified oligonucleotides, the half life of phosphorothioate oligonucleotides in many experimental animals is less than one hour (S. Agrawal, *et al.*, 88 Proc. Natl. Acad. Sci. USA 7595-7599 (1991); P. Iverson, 6

Anticancer Drug Design 531-538 (1991)) such that delivery of an efficacious dose of the oligonucleotide may be problematical. Methylphosphonate oligonucleotides do not activate RNase H enzyme activity (L.J. Maher, *et al.*, 245 Science 725-730 (1989); P.S. Miller, *in* J.S. Cohen (Ed.), Oligodeoxynucleotides: Antisense Inhibitors of Gene Expression, CRC Press: Boca Raton, p, 79 (1989)) and are eliminated rapidly (T.L. Chen, *et al.*, 18 Drug Metab. Dispos. Biol. Fate. Chem. 815 (1990)).

Another factor influencing the delivery of synthetic oligonucleotides into cells is the low permeability of the cell membrane to such compounds. Low permeability may effectively preclude adequate uptake for sustained *in vivo* activity. Consequently, many oligonucleotides must be delivered to the cell by carrier systems such as liposomes or molecular complexing agents. However, the relatively short circulating time of liposomes, lack of targeting specificity, and the need for repeat exposure severely limits the usefulness of synthetic oligonucleotide delivery schemes. These difficulties can be avoided, however, if production of the desired single-stranded oligonucleotide takes place within the cell (*in vivo*).

Nevertheless, so far as is known, no method for producing single-stranded nucleic acid in eukaryotic cells is available which overcomes all of these limitations and disadvantages. It is, therefore, an object of the present invention to provide a method which overcomes these limitations and disadvantages, and in more detail, it is an object of the present invention to provide a DNA construct which directs the synthesis of ssDNA *in vivo*. It is another object of the present invention to provide a method which reduces and/or eliminates the problems due to toxicity, specificity, and RNAse H inhibition or inactivation of prior systems by a continuous process which synthesizes naturally-occurring phosphodiesterase single-stranded DNA within the target cell.

Another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA of any desired nucleotide sequence within eukaryotic cells without undesirable intervening or flanking nucleotide bases.

Another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA *in vivo* in a manner which isolates or presents a sequence of interest to the desired target without interference from intervening or flanking nucleotide bases.

Another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA of any nucleotide sequence *in vivo* which can be used for (but is not limited to) binding to mRNA in an anti-sense fashion to down regulate a gene product or a viral gene product of interest.

Another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA designed in such a way as to favor binding to duplex (native DNA) to form triplex structures which may interfere with normal gene transcription and regulation.

Another object of the present invention is to produce ssDNA within eukaryotic cells for the purpose of disrupting one or more of the many highly regulated cell functions. For instance, the ssDNA tails of telomeric repeats may be altered by the production of ssDNA which has identical or complimentary nucleotide base composition to the sequence of the native DNA in the telomeric repeats or other regulatory sequence.

Another object of the present invention to provide a method, and a DNA construct, for producing ssDNA *in vivo* which is designed so that the nucleotide sequence is recognized by a cellular protein of interest and subsequently binds to and inhibits a specific cellular function, for instance, by binding to proteins which recognize nucleic acid sequences.

Yet another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA into which secondary structures are designed so that the ssDNA oligonucleotides bind to and/or otherwise inhibit or activate various cellular functions which rely on nucleic acid protein interaction such as transcription, translation, and DNA replication.

Another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA *in vivo* for site-directed mutagenesis or gene knockout for therapeutic applications.

Another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA of precisely defined nucleotide composition which favors site-specific insertion into a genome for therapeutic purposes.

Yet another object of the present invention is to provide a method, and a DNA construct, for producing ssDNA that is complimentary to any endogenous nucleic acid sequence target.

SWB A2

Still another object of the present invention is to provide a unique ss-cDNA "stem-loop" structure that is designed so that the stem is comprised of an inverted tandem repeat that folds back on itself to form a stable, double-stranded stem that terminates reverse transcriptase products, thereby eliminating undesirable 5' sequence portions of the intermediate RNA transcript needed for ssDNA formation *in vivo*.

Another object of the present invention is to provide a method, and pharmacologically acceptable compositions, for delivery of antisense sequences to the nucleus of target cells in a manner which produces a therapeutic effect.

Another object of the present invention is to provide a method of using multiple combinations of enhancers and/or promoters, which may or may not containe inducible control elements, to drive the expression of the enzymes needed for *in vivo* synthesis of single-stranded DNA from an RNA transcript.

Another object of the present invention is to provide a method of using stem-loop intermediates which are designed such that reverse transcription of an mRNA to make ss-cDNA is terminated before transcription of the entire mRNA transcript.

Another object of the present invention is to provide a method of using a stem-loop intermediate designed such that variable amounts of premature termination and/or read-through of ss-cDNA occurs in the cell for regulating production of a series of nucleic acid sequences of interest.

Another object of the present invention is to provide a method for the production and regulation of *in vivo* ss-cDNA by secondary folding of mRNA directed transcripts.

This listing of the objects of the present invention is not intended to be a list of all of the objects of the invention. There are a vast number of other cellular functions which are mediated by the cellular genome which, in the interest of brevity and practicality, are not mentioned here and which are amenable to regulation by *in vivo* production of ssDNA. For instance, exonucleases digest ssDNA much more aggressively than double-stranded DNA (dsDNA). Consequently, another object of the present invention is to provide a ssDNA construct, and a method of producing that construct *in vivo*, which is not as susceptible to degradation by native exonucleases in the cell as ssDNA. It can be seen from this illustration that this list of some of the objects of the present invention is provided for purposes of exemplification and is not intended to limit the scope of the present invention.

SUMMARY OF THE INVENTION

These objects, and the many others which will be made apparent to those skilled in the art by the following description of the presently preferred embodiments of the invention, are achieved by providing a gene encoding a polyprotein which contains RNA-dependent DNA polymerase/reverse transcriptase co-joined with RNAse H, a cassette, or nucleic acid construct, comprised of a sequence of interest flanked by inverted tandem repeats, and the system for expressing the cassette in a host cell. The cassette also preferably includes a second sequence of interest 3' to the inverted repeats for premature termination of transcription reactions and either constitutive or inducible eukaryotic promoter(s)/enhancer(s) for the RNA-dependent DNA polymerase. The cassette may also include a restriction endonuclease gene and an appropriate promoter for the particular restriction endonuclease being utilized. The invention also contemplates that the cassette is incorporated into a plasmid and that the plasmid is incorporated into a suitable host cell.

In another aspect, the present invention comprises a method of producing single-stranded DNA *in vivo* comprising the steps of transcription of a cassette comprising an RNA-dependent DNA polymerase gene and a sequence of interest in a eukaryotic cell and converting the mRNA transcript of the sequence of interest to cDNA with the polymerase produced by the RNA-dependent DNA polymerase gene with simultaneous digestion of the mRNA component template with an RNase H expressed enzyme. The sequence of interest also includes an inverted tandem repeat. The cassette may also include a restriction endonuclease gene which, when transcribed and translated, produces a restriction endonuclease which linearizes the transcript of the sequence of interest by cutting the ss-DNA at a restriction endonuclease site formed when the inverted tandem repeat causes the cDNA transcript to form a stem-loop intermediate.

In another aspect, the present invention comprises a method for producing a single-stranded oligonucleotide in a target cell. In one embodiment, this method is intended to deliver an antisense sequence. In other embodiments, the method is used to deliver triplex-forming sequences or sequences which are recognized and bound by specific DNA-binding proteins, or other nucleic acids and/or proteins which function in cellular metabolism and/or replication.

The method comprises the encoding of the oligonucleotide into a complementary sequence of interest in a cassette which includes a gene encoding for an RNA-dependent DNA polymerase which preferably includes an RNase H gene and an inducible or constitutive eukaryotic promoter/enhancer appropriate for that polymerase/RNase H gene. The cassette includes a gene encoding a restriction endonuclease (RE) and, in the preferred embodiment, an appropriate promoter/enhancer for that RE gene. The cassette further comprises an inverted tandem repeat and, when assimilated into the target cell, the cassette (including the sequence of interest and the inverted tandem repeats) is transcribed by the cell under the control of the promoter(s)/enhancer(s). The normal function of the target cell causes the resulting mRNA transcript of the polymerase and RE genes to be translated, providing all that is needed for production of ss-DNA from the mRNA transcript of the sequence of interest. Specifically, the RNA-dependent DNA polymerase produced from the cassette converts the mRNA transcript of the sequence of interest and inverted tandem repeats to ss-cDNA, the ss-cDNA forms a stem-loop intermediate as the nucleotide bases comprising the inverted tandem repeats pair up by Watson-Crick base pairing, and the restriction endonuclease produced from the RE gene digests the double-stranded portion of the stem-loop intermediate to "free" the single stranded DNA oligonucleotide from the loop portion of the stem-loop intermediate.

The present invention also comprises methods and compositions for producing human telomere repeat sequence $(TTAGGG)_n$, and the complimentary sequence $(CCCTAA)_n$, in eukaryotic cells in the manner described in the preceding paragraphs.

In another aspect, the present invention comprises a method and a vector for producing ss-cDNA in eukaryotic cells, a method of producing ss-cDNA in eukaryotic cells by enzymatic reactions of reverse transcriptase/RNAse H of any source or origin which have been cloned into eukaryotic cells, and a method of producing the reverse transcriptase and RNAse H enzymes in eukaryotic cells.

In another aspect, the present invention achieves the intended results by providing plasmids including the gene sequences, and a method of using RNAse H activity expressed in eukaryotic cells, to digest RNA-cDNA hybrids and a method of using reverse transcriptase and RNAse H activities expressed in eukaryotic cells to produce ssDNA by digesting such hybrids.

The present invention also provides a method of using multiple combinations of enhancers and promoters, which may or may not include inducible control elements, to drive the expression of the necessary enzymes and RNA transcript for ssDNA production *in vivo*.

Also provided are methods and plasmids for producing stem-loop intermediates *in vivo* which are capable of variable amounts of read through of a stem-loop mRNA intermediate or early termination as the ss-cDNA is being synthesized for regulating production of the ss-cDNA from one or more sequences of interest. In other words, the present invention provides methods for production and regulation of in vivo ss-cDNA by secondary folding of mRNA directed transcripts.

Multiple systems can be used to deliver the cassette to the target cell to direct the synthesis of ssDNA within the cell, including plasmid or plasmid-based vector systems or viral based vector systems, and these systems are adapted for that purpose in accordance with standard delivery techniques currently known to the skilled practitioner. These systems include, but are not limited to, viral based systems such as adenovirus, adenoassociated virus, retroviral vectors, and conjugate vectors using double stranded plasma DNA based transfection systems. All such systems are contemplated by the present invention. Once inside the cell, the cassette is transcribed in the normal course of cell metabolism, producing an mRNA transcript of the sequence of interest that is then converted to cDNA by the reverse transcriptase which is likewise produced by the cell from the reverse transcriptase/RNase H gene included in the cassette under the control of the promoter.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic map of the plasmid pssDNA-Express A constructed in accordance with the present invention.

Figures 2A, 2B, and 2C represent a schematic map of the plasmid pss-DNA-Express B constructed in accordance with the present invention, an enlarged portion of the plasmid pss-DNA-Express B, and the sequence of the insert region of plasmid pss-DNA-Express B, respectively.

Figure 3 represents the sequence of the insert regions between the *Apa* I and *Nhe* I sites of the plasmids pTest and pTelo constructed from the plasmid pss-DNA-Express B in accordance with the present invention.

Figures 4A and 4B are gels showing the results of PCR reverse transcriptase assays (J. Silver, *et al.*, 21 Nucleic Acids Res. 3593-3594 (1993)) on RNA/ssDNA extracts from HeLa cell lines transformed with the plasmid pTest constructed in accordance with the present invention. Lanes, Fig. 4A: (1) size markers, (2) untransformed HeLa cells, (3) HeLa cells transformed with the vector pcDNA3.1Zeo$^+$, (4) HeLa cells transformed with pssXA, (5) positive control containing 2 milliunits of MoMuLV reverse transcriptase, and (6) negative control containing no protein extract. The arrow indicates the 150 bp amplification product. Lanes, Fig. 4B: (1) clone A12, (2) B8, (3) B12, (4) D7, (5) positive control containing 2 milliunits of MoMuLV reverse transcriptase, and (6) negative control containing no protein extract.

Figure 5 are gels showing PCR amplification of single-stranded DNA from extracts of HeLa cell lines transformed with the plasmid pTest constructed in accordance with the present invention. RNA/ssDNA preparations were used as templates in PCR reactions using primers specific to the expected ssDNA product. Left panel, lanes 1-3: total RNA/ssDNA PCR template isolated from stably transformed colony A12 (HeLa cell line stable transfected with plasmid pssDNA-Express-A) (1) used with no prior treatment, (2) treated with S1 nuclease, (3) treated with RNAse A. Lane (4) total RNA/ssDNA from a colony stably transformed with the pcDNA3.1Zeo$^+$ vector alone. Right panel, lanes 1-3: total RNA/ssDNA PCR template isolated from stably transformed colony B12 (HeLa cell line stably transfected with plasmid pssDNA-Express-A) (1) used with no prior treatment, (2) treated with S1 nuclease, (3) treated with RNAse A. Lane (4) total RNA/ssDNA from a colony stably transformed with the pcDNA3.1Zeo$^+$ vector alone. Lane (5), positive control template, plasmid pTest.

Figure 6 are gels showing PCR amplification of single-stranded DNA from extracts of cells transformed with the plasmid pTelo constructed in accordance with the present invention. RNA/ssDNA was harvested from transfected cell cultures at 36 hrs post transfection and used as templates for PCR reactions using primers specific to the expected ssDNA product. 25 bp lane marker used. Lanes 1-5 with isolation from HeLa cell line A12, (1) total RNA/ssDNA fraction, (2) total RNA treated with S1 nuclease, (3) total RNA treated with RNAse A, (4) negative control, (5) positive control telomere DNA plasmid. Lanes 6-9 RNA isolated from B12 cell line, (6) total RNA fraction, (7) total RNA treated with S1 nuclease, (8) total RNA treated with RNAse A, (9) negative

control, (10) positive control telomere DNA plasmid. Lanes 11-12 repeat negative controls. 8% acrylamide gel at 45 V for 20 mins.

Figure 7 is a gel showing premature truncation of ss-cDNA transcription and read through products by *in vitro* reverse transcriptase reaction. Lanes 1-5 represent plasmid constructs with stem-loop structures constructed in accordance with the present invention and various sequences of interest cloned within the loop region. RNA template was produced by *in vitro* transcription with T7 RNA polymerase under standard conditions, then treated with DNAse to remove plasmid template. Phenol/chloroform extracted RNA pools were then reverse transcribed with mouse Moloney reverse transcriptase, treated with RNAse A for 15 mins. and resolved on a 6% acrylamide gel at 45 V for 30 mins. 25 bp marker at borders. (1) 10 µl RT reaction from original plasmid pssDNA-Express-B constructed in accordance with the present invention, (2) 10 µl RT reaction from original plasmid pTest constructed in accordance with the present invention, (3) 10 µl RT reaction from original plasmid pTelo constructed in accordance with the present invention, (4) negative control plasmid pcDNA-Zeo (Invitrogen), (5) 3 µl repeat pssDNA-Express-B.

Figure 8 is a schematic representation of ssDNA production and premature termination of the ssDNA transcript *in vivo*.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In this description of the preferred embodiments of the present invention, methods and nucleic acid constructs are described for use in producing single-stranded deoxyribonucleic acid (ss-cDNA) oligonucleotides of virtually any predefined or desired nucleotide base composition *in vivo* in eukaryotic cells with or without flanking nucleotide sequences. Methods and constructs are described which use biological rather than the *in vitro*, or artificial chemical synthesis of ss-cDNA of desired nucleotide base composition. Because biological, i.e., enzymatic reactions, are used in these methods they are applicable to any *in vivo* system.

A vector (as used herein, the term "vector" refers to a plasmid or modified viral construct used to deliver and manipulate DNA segments of interest) set was designed to produce any DNA sequence as a ss-cDNA molecule, free of most contiguous vector sequence, within mammalian cells. The plasmids contain all the necessary enzymatic functions and signaling instructions to allow the host cell to produce ss-cDNA. The

other component of the present invention is an RNA transcript, driven by an eukaryotic promoter just as the above-described enzymes are driven by eukaryotic promoters, which is used as a template by the enzymes described above to direct the synthesis of any desired single-stranded DNA sequence (a "sequence of interest").

In brief, in a first aspect, the components utilized to produce ss-DNA *in vivo* in accordance with the present invention include a RNA dependent DNA polymerase (reverse transcriptase) gene, a primer binding site (PBS) for the reverse transcriptase, a restriction endonuclease (RE), a sequence of interest located between inverted repeats (IR) or 3' to the inverted repeats, a primer binding site located 3' to the cassette including the sequence of interest and the inverted repeats, and the functions and signaling instructions for transcription of these components *in vivo*. In a second aspect, the components utilized to produce ss-DNA *in vivo* in accordance with the present invention include a reverse transcriptase gene, a sequence of interest located between inverted repeats or 3' to the inverted repeats, and a primer binding site located 3' to the cassette including the sequence of interest and the inverted repeats. The invention also contemplates a second sequence of interest located between the cassette (including a sequence of interest and the inverted repeats) and the primer binding site, and the functions and signaling instructions for transcription of these components *in vivo*.

In the presently preferred embodiment, these components are inserted into a host cell by transfection of the cells with two plasmids, designated pssDNA-Express-A and pssDNA-Express-B, each plasmid being designed and constructed to include certain of these components. One plasmid (pssDNA-Express-B in the preferred embodiment) encodes the sequence of interest, nested within flanking sequences that include the inverted repeats, and the primary binding site that provide the post-transcriptional processing signals that mediate the conversion of the mRNA into single-stranded DNA. The B plasmid also includes the second sequence of interest when this second aspect of the invention is utilized. Activities required for processing the primary gene product of the B plasmid into single-stranded DNA, with the removal of vector sequences and processing signals, specifically the reverse transcriptase/RNAse H, and restriction endonuclease, are expressed from the A plasmid (pssDNA-Express-A in the preferred embodiment). The single-stranded DNA sequence that is released by interaction of the transcriptional products of these components *in vivo* is free to bind intracellular targets

such as mRNA species and DNA promoters in antisense and triplex strategies. In particular, in addition to a control sequence, the data set out herein describe expression of a single-stranded telomeric repeat.

In more detail, in the preferred embodiment described herein, the B plasmid pssDNA-Express-B includes cloning sites (*Not* I sites were utilized in the preferred embodiment of the B plasmid described herein) between which any DNA sequence of interest may be placed (as noted above, in the examples described herein, the sequences include a test, or control sequence and a telomeric repeat). Flanking the cloning sites are signals directing the processing of the primary mRNA transcript, produced from a promoter (a CMV promoter was utilized in the preferred B plasmid described herein), into the desired single-stranded DNA. After cloning of the desired sequence of interest into the B plasmid, the A and B plasmids are co-transfected into a cell line of choice for constitutive expression of ssDNA. This processing proceeds in three steps following transcription of the single-stranded DNA region (i.e., sequence of interest, inverted repeats, and PBS):

(1) reverse transcription of the B plasmid RNA transcript by a reverse transcriptase, which in the preferred embodiment described herein is expressed by the A plasmid (in the preferred embodiment described herein, the reverse transcriptase is Moloney mouse leukemia virus (MoMuLV) reverse transcriptase), proceeding from a primer site lying 3' to the cassette comprising the sequence of interest and the inverted repeats;

(2) RNAse H digestion of the resulting heteroduplex, either by the RNAse H activity of the reverse transcriptase polyprotein or by endogenous RNAse H activity, to release the single-stranded DNA precursor from its RNA complement; and

(3) In the case of the second aspect of the invention, premature termination of the cDNA transcript by formation of the stem-loop secondary structure by the self-complementary inverted tandem repeats.

Those skilled in the art will recognize from this disclosure that the particular cloning sites flanking the sequence of interest, the particular reverse transcriptase, restriction endonuclease, promoter, primer binding site, and all the other components of the construct described herein may be chosen depending upon the particular sequence of

interest and/or system in which the ssDNA is to be expressed. It will also be recognized that these components may be incorporated into a single plasmid rather than the two plasmids described herein.

Regarding the RNA-dependent DNA polymerase, or reverse transcriptase (RT) gene which is the first component of the present invention, as noted above, the reverse transcriptase/RNAse H gene from Moloney murine leukemia virus was used to advantage in the examples described herein. The reverse transcriptase/RNAse H gene from the human immunodeficiency virus (HIV) was also tested. Many other retroviral reverse transcriptase/RNAse H genes may be used to advantage in connection with the present invention, it being preferred that the reverse transcriptase/RNase H gene be a reverse transcriptase/RNase H gene that is regulated by an appropriate upstream eukaryotic promoter/enhancer such as the CMV or RSV promoter for expression in human cells.

Many RNA-dependent DNA polymerase/reverse transcriptase genes are known which are suitable for use in connection with the present invention including those from retroviruses, strains of hepatitis B, hepatitis C, bacterial retron elements, and retrons isolated from various yeast and bacterial species. As found in nature, these RNA-dependent DNA polymerases usually have an associated RNase H component enzyme within the same coding transcript. However, the present invention does not require the naturally-occurring RNase H gene for a particular reverse transcriptase. In other words, those skilled in the art will recognize from this disclosure that various combinations of reverse transcriptase and RNase H genes can be spliced together for use in connection with the present invention to fulfill this function and that modifications and/or hybrid versions of these two enzyme systems are available and/or known to those skilled in the art which will function in the intended manner. Those skilled in the art will also recognize that the target cell may itself have sufficient endogenous RNase H to fulfill this function. Similarly, those skilled in the art will recognize that the target cell may itself have sufficient endogenous reverse transcriptase from, for instance, prior retroviral infection, to fulfill this function.

Those skilled in the art who have the benefit of this disclosure will also recognize that a number of tissue-specific or wide spectrum promoters/enhancers, or combinations of promoters/enhancers other than those described herein may also be used to advantage

to control the reverse transcriptase/RNAse H gene, the RE gene, and the sequence of interest. Although a list of all available promoters/enhancers is not needed to exemplify the invention, as noted above, the promoters/enhancers may be constitutive or inducible and may include the CMV or RSV promoters/enhancers listed here and many other viral or mammalian promoters. Representative promoters/enhancers which are appropriate for use in connection with the present invention may include, but are not limited to, HSVtk (S.L. McKnight, *et al.*, 217 Science 316 (1982)), human ß-globulin promoter (R. Breathnach, *et al.*, 50 Ann. Rev. of Biochem. 349 (1981)), ß-actin (T. Kawamoto, *et al.*, 8 Mol. Cell Biol. 267 (1988)), rat growth hormone (P.R. Larsen, *et al.*, 83 Proc. Natl. Acad. Sci. U.S.A. 8283 (1986)), MMTV (A.L. Huang, *et al.*, 27 Cell 245 (1981)), adenovirus 5 E2 (M.J. Imperiale, *et al.*, 4 Mol. Cell. Biol. 875 (1984)), SV40 (P. Angel, *et al.*, 49 Cell 729 (1987)), α-2-macroglobulin (D. Kunz, *et al.*, 17 Nucl. Acids Res. 1121 (1989)), MHC class I gene H-2kb (M.A. Blanar, *et al.*, 8 EMBO J. 1139 (1989)), and thyroid stimulating hormone (V.K. Chatterjee, *et al.*, 86 Proc. Natl. Acad. Sci. U.S.A. 9114 (1989)).

The reverse transcriptase/RNase H gene also preferably includes a downstream polyadenylation signal sequence so that the mRNA produced from the reverse transcriptase/RNase H gene includes a 3′ poly(A) tail for mRNA stability. As known to those skilled in the art, multiple poly(A) tails are available and are routinely used for production of expressed eukaryotic genes.

The reverse transcriptase produced in the cell synthesizes a complementary DNA (cDNA) using as the template the genetic element including the sequence of interest described below. The RNase H activity of the reverse transcriptase degrades the mRNA template component of the RNA/cDNA hybrid to produce ss-DNA *in vivo*.

The gene encoding the restriction endonuclease may be any of several genes which encode for restriction endonucleases, and preferably those that are controlled by one or more constitutive or inducible wide spectrum and/or tissue-specific promoters/enhancers such as those listed above. The particular restriction endonucleases tested were MboII and FokI, Mbo II being preferred, but those skilled in the art who have the benefit of this disclosure that any restriction endonuclease (type I, II, IIS, or III) site may be included in the inverted tandem repeat. These enzymes "clip" or digest the stem

of the stem-loop intermediate described below to linearize the sequence of interest as single-stranded DNA.

Although expression of this second enzyme may be regulated by an appropriate constitutive or inducible promoter/enhancer located upstream from the restriction endonuclease gene such as the CMV or RSV promoter for expression in human cells, in the preferred plasmid pssDNA-Express-A, the RE gene (MboII) is linked to the RT-RNAse H polypeptide. Those skilled in the art who have the benefit of this disclosure will also recognize that several other promoters/enhancers such as those listed above may be used to advantage to control the RE gene just as various promoters/enhancers are available for the reverse transcriptase/RNase H gene as discussed above. The RE gene also preferably includes a downstream polyadenylation signal sequence so that the mRNA transcript from the restriction endonuclease gene will have a 3′ poly(A) tail.

The third component of the invention encodes a nucleic acid sequence that is referred to herein as a cassette which provides the template for synthesis of ss-cDNA in target cells. It is this element which includes the sequence of interest and the inverted tandem repeats. As is the case for the above genes, this genetic element is preferably regulated by an appropriate wide spectrum or tissue-specific promoter/enhancer, such as the CMV promoter, or combination of promoters/enhancers, located upstream of the genetic element. Also as was the case for the above genes, the promoter/enhancer can either be constitutive or inducible promoter. Those skilled in the art who have the benefit of this disclosure will recognize that, as noted above, a number of other eukaryotic promoters may be used to advantage to control expression of the sequence of interest including SV-40, RSV (non-cell type specific) or GFAP (tissue specific).

For expression in eukaryotic cells, the cassette also includes a downstream polyadenylation signal sequence so that the mRNA produced by the sequence of interest has a poly(A) tail. Between the 3′ inverted tandem repeat and the polyadenylation signal, the genetic element contains a primer-binding site (PBS) for initiation of priming for cDNA synthesis. The PBS is a sequence which is complementary to a transfer RNA (tRNA) which is resident within the eukaryotic target cell. In the case of the mouse Maloney reverse transcriptase described herein as being utilized in conjunction with the present invention, the PBS takes advantage of the lysine tRNA. The PBS utilized in connection with the presently preferred embodiment of the invention that is described

herein was taken from the actual 18 nucleotide sequence region of mouse Moloney virus. *See* 293 Nature 81. In the case of the reverse transcriptase gene from human immunodeficiency virus also described above, the PBS used was taken from the nucleotide sequence of HIV. Y. Li, *et al.*, 66 J. Virology 6587-6600 (1992). In short, any PBS that is matched to the reverse transcriptase which is utilized in connection with the method of the present invention is utilized for this purpose.

In the second aspect of the invention, a second sequence of interest is inserted between the primer binding site and the 3′ inverted tandem repeat. The second sequence of interest is likewise reverse transcribed from the PBS. Depending upon the stabilility of the stem loop secondary structure, reverse transcription may be terminated at the stem loop formed in the mRNA transcript of the cassette such that only the second sequence of interest is produced as ssDNA in the host cell or the second sequence of interest is reverse transcribed along with the inverted repeat and the first sequence of interest. In a third aspect, either of the first or second sequences of interest may also be included in the cassette with multiple additional sequences of interest, each with its corresponding promoter/enhancer, polyadenylation signal, and PBS (inside or outside of the inverted tandem repeats), and PBS, can be utilized for producing ssDNA *in vivo*, for instance, delivering anti-sense sequences to the target cell.

The inverted tandem repeats cause the ss-cDNA to fold back upon itself to form the stem of a stem-loop structure in the manner described in my co-pending application Serial No. 08/877,251, the specification of which is hereby incorporated into this application in its entirety by this specific reference. The folding of the ss-cDNA in this fashion occurs after the sequence of interest and its flanking inverted repeats are transcribed in the cell and after the reverse transcriptase/RNase H (produced by transcription of those genes which are also included) produce the ss-cDNA sequence of interest from the mRNA transcript in the cell. The stem comprises one or more restriction endonuclease site(s) which is cut by the restriction endonuclease produced from the restriction endonuclease gene likewise coded for by the corresponding component of the invention as described in co-pending application Serial No. 08/877,251. The ss-cDNA which is produced is transcribed with the encoded 5' and 3' regions flanking the stem (made up of the inverted repeats) and the loop (containing the sequence of interest). This stem structure is comprised of double stranded, anti-parallel

DNA and is designed to contain one or more restriction endonuclease recognition sites within the double stranded portion, i.e., the inverted repeats. In this manner, the stem is cut (also termed digested or cleaved) by any of the many restriction endonuclease enzymes which recognize the cut site designed into the stem. The loop portion of the ss-cDNA, which does not form any apparent duplex DNA, is immune to restriction endonuclease activity since restriction endonucleases recognize only double stranded DNA as a target substrate.

It will be recognized by those skilled in the art that the restriction endonuclease site(s) need not be designed into the inverted repeats which form the stem of the stem-loop intermediate if the second aspect of the present invention is being utilized. In other words, if it is desired to produce ssDNA from a second sequence of interest located between the primer binding site and the inverted repeats, with transcription of the cassette to terminate at the stem formed by the inverted repeats, there is no need for a restriction endonuclease site in the stem. Another option is to design the inverted repeats to contain eukaryotic, prokaryotic, or viral protein DNA binding sites, which can act to competitively titer out selected cellular proteins. Combinations of restriction sites or other genetic elements may be included in the inverted tandem repeats depending on the base pair composition chosen for the construction of inverted repeats such that linear or precisely cut stem-loop intermediate forms of ss-DNA are produced. It is generally preferred to use synthetically constructed genetic elements in the inverted tandem repeats since it is unlikely that a naturally occurring inverted repeat would have the properly aligned restriction sites.

When the components of the present invention are incorporated into a vector, it is preferred that a shuttle vector is used so that the components can be amplified on prokaryotic systems and then subsequently expressed in eukaryotic systems. It is also preferred that the vector include a selective marker gene. For example, it is preferred that amplification of the vector be accomplished in prokaryotic systems which preferably include ampicillin, kanamycin, or tetracycline resistance genes for positive selection in those systems. For expression in eukaryotic systems, multiple selection strategies may be used including, but not limited to, resistance markers for Zeocin, resistance to G418, or phenotypic selection markers such as *B-gal* or green fluorescence protein.

Incorporation of these components into an appropriate vector allows two convenient methods for removing predetermined vector sequences after the production of ssDNA. In the first method, the loop portion of the ssDNA that is produced is comprised of the nucleotide sequence of interest and after digestion with the restriction endonuclease, the loop is released as linearized, single-stranded cDNA without any flanking sequences. In the second method in which the cassette is reverse transcribed from the PBS and a second sequence of interest is included in the cassette 3′ to the inverted tandem repeat, reverse transcription is terminated at the stem of the stem-loop structure such that the resulting ssDNA is produced without flanking sequences. If it is desired to utilize the second method, the cassette is designed with inverted repeats which form a more stable stem. By designing the cassette with inverted repeats that form a stem that is easily denatured, reverse transcription proceeds right on through the second sequence of interest (if it is even designed into the cassette). A stem which is intermediate in stability will allow production of both the first and second sequences of interest.

In more detail, and using the specific genes and other constructs described below to exemplify the mode of action of the methods and constructs of the present invention, when the two A and B plasmids are co-transfected into a suitable host cell, which can be any eukaryotic cell, single-stranded DNA having the sequence of interest is produced in the cell. The sequence of interest is placed in the B plasmid, pssDNA-Express-B in the example set out below, so that the sequence of interest is transcribed from the cytomegalovirus promoter, which is located just upstream of the sequence of interest, and terminates with the BGH polyA signal located just downstream of the sequence of interest. Reverse transcriptase, expressed from the A plasmid, pssDNA-Express-A in the examples set out below, reverse transcribes the primary transcript, using host t-RNApro as a primer, from the position of the MoMuLV reverse transcriptase promoter, producing a DNA-RNA duplex molecule. Endogenous RNAse H activity, or RNAse H produced from the A plasmid, degrades the RNA stand, releasing a single-+stranded DNA as shown in Fig. 8. This ssDNA sequence contains inverted, complementary repeats which form a stem-loop structure with the restriction endonuclease site, specifically, an Mbo II recognition site (GAAGA) in the stem region. Mbo II, also expressed from the A plasmid pBK-RSV-RT/Mbo-L (also referred to herein as pssDNA-Express-A), cuts the

double-stranded stem outside of its recognition sequence, releasing the loop region as a single-stranded DNA of sequence and length defined by the cut sites and the particular sequence of interest. The reverse transcriptase and Mbo II activities are expressed in a single protein chain encoded by a transcript produced from the RSV promoter of the A plasmid and separated by a short linker region rich in proline.

Premature termination of the reverse transcriptase cDNA transcript at the 3' aspect of the stem structure was discovered when the vector pNM-new-link described below was used in the *in vitro* experiments described below with a 29 base pair stem structure, allows a unique construct and method for limiting the intervening vector sequences contained with an *in vivo*-produced ss-cDNA. Specifically, a new construct having the sequence of interest located 3' in relation to the stem-loop structure and 5' in relation to the PBS as shown in Fig. 8 is described and used for producing ssDNA *in vivo*.

Premature reaction termination by secondary structure of the mRNA transcript is used as an alternate control mechanism to prduce various size products from *in vivo* cDNA producing reactions. Limiting the extent of 3′ read through of the ss-cDNA into more proximal 5′ mRNA substrate has the effect of reducing the amount of or eliminating unwanted 5′ mRNA transcript sequences which could interfere with the desired function of the ss-cDNA product (i.e., for antisense or triplex formation). Those skilled in the art will recognize from this disclosure that the placement of such stem-loop intermediates in any desired loaction within the mRNA transcript allows modification of the ss-cDNA produced by reaction termination as a result of the secondary structure imposed on the transcript. It will also be apparent from the results of the experiment reported below and from Fig. 7 of the present application that a mixed population of ss-cDNA (i.e., the truncated transcript at the position between 3′ PBS and 3′ aspect of the inverted repeat and the read through product) is produced by transformation of host cells with the pNM-New-Link plasmid described herein and that the multiple trasncripts produced may be assigned different functions in the cell. For example, if different antisense sequences are assigned to different positions in the mRNA transcript (i.e., one to the loop portion of the stem-loop secondary structure as the first sequence of interest and one to the position between the PBS and the stem as the second sequence of interest) and produced in different proportions depending upon the stability of the stem of the

stem-loop structure, various levels of control are possible within the cell and at the transcriptional level.

In addition to production of multiple ss-cDNA sequences *in vivo*, the present invention contemplates the purposeful introduction of cellular factors such as single and double stranded binding proteins into the host cells to recognize the double-stranded stem portion or single stranded/double stranded junction of the stem-loop transcript to add to or reduce the stability of this secondary structure. The use of these factors allows control over the amount of premature termination or read through products produced by the cell. Quantitative changes, as well as changes ss-cDNA length, are orchestrated by manipulation of the stem-loop structure within the mRNA transcript in the cell. Thus, the secondary structure of the template mRNA is utilized to produce and control the production of ssDNA *in vivo*.

It will also be evident to those skilled in the art from this description that the intact stem-loop ss-cDNA structure can function similarly in many applications as the linearized ss-cDNA form. Consequently, the cassette is also used to advantage without the restriction endonuclease gene and associated regulatory elements and/or with a sequence of interest which lacks the corresponding restriction endonuclease site.

It will also be evident to those skilled in the art from this description of the preferred embodiments of the present invention that a cassette can be made which encodes a ss-cDNA that has a "trimmed" stem-loop structure. The restriction endonuclease sites encoded in the inverted repeats flanking the sequence of interest are designed such that the stem portion (after duplex formation) is digested with the corresponding restriction endonuclease so as to cut the dsDNA comprising the stem in a way that removes a portion of the stem and the associated flanking sequences yet leaves sufficient duplex DNA that the transcript retains the above-described stem-loop structure. Such a ss-cDNA structure may be more resistant to intracellular nucleases by retaining the "ends" of a ssDNA in double stranded form.

It will also be evident to those skilled in the art from this description of the preferred embodiments of the invention that the stem (duplex DNA) can be designed to contain a predetermined sequence that is recognized and bound by specific DNA-binding proteins. Among other uses, such a stem structure is used in the cell as a competitor to

titer out a selected protein(s). For example, production of the ss-cDNA stem-loop of the present invention in a cell that contains a binding site for a selected positive transcription factor such as E1a which acts to "bind up" the factor, preventing it from binding a particular promoter and thus inhibiting the expression of a particular deleterious gene. Any desired nucleotide sequence can be inserted into the genetic element which encodes the 'loop' portion to ultimately carry out a desired function, e.g., antisense binding, down regulation of a gene, and so on as herein described.

In another aspect which will be recognized by those skilled in the art, the present invention is used to construct complex secondary ssDNA structures which confer biologic reactions on the cDNA transcript produced in accordance with the present invention based on conformational secondary structure folding such as DNA enzymes (S.W. Santoro and G.F. Joyce, 94 Proc. Natl. Acad. Sci. USA 4262-4266 (1997)).

. Such secondary structure can be engineered to serve any of several functions. For instance, the sequence of interest may include (but is not limited to) a sequence which is incorporated into the loop portion of the single-stranded cDNA transcript which forms so-called "clover leaf" or "crucible" like structures such as those found in the long terminal repeats of adeno-associated virus or in retrotransposons. Under correct circumstances, such structure is integrated in site-specific manner into the host genome.

Because the cassette of the present invention is adaptable for incorporation into multiple commercially available delivery vectors for mammalian and human therapeutic purposes, multiple delivery routes are feasible depending upon the vector chosen for a particular target cell. Such systems include intravenous, intramuscular, and subcutaneous injection, as well as direct intra-tumoral and intra-cavitary injections. The cassette, when inserted into the vector of choice can also be administered through topical, transmucosal, rectal, oral, or inhalation-type methods of delivery.

The cassette of the present invention is advantageously employed to deliver anti-sense, triplex, or any other single-stranded nucleotide sequence of interest, using known digestion and ligations techniques to splice the particular sequence of interest into the cassette between the inverted tandem repeats. Those skilled in the art who have the benefit of this disclosure will also recognize that the above-described signals used for expression within eukaryotic cells may be modified in ways known in the art depending

upon the particular sequence of interest. The most likely change is to change the promoter so as to confer advantageous expression characteristics on the cassette in the system in which it is desired to express the sequence of interest. There are so many possible promoters and other signals, and they are so dependent on the particular target cell for which the sequence of interest has been selected, that it is impossible to list all the potential enhancers, inducible and constitutive promoter systems, and/or poly(A) tailing systems which may be preferred for a particular target cell and sequence of interest.

In one particularly preferred embodiment, the present invention takes the form of a kit comprised of a plasmid having the above-described RNA-dependent DNA polymerase and restriction endonuclease genes cloned therein as well as a cloning site for into which the user of the kit ligates a particular sequence of interest. The cloning site into which the sequence of interest is ligated is located between the above-described inverted tandem repeats. The resulting plasmid is then lyophilized or otherwise preserved for packaging and shipping to the user. The kit preferably also includes the restriction endonuclease for the cloning site into which the sequence of interest is to be cloned and a map of the plasmid along with suitable buffers for ligating the sequence of interest into the cloning site.

Except where otherwise indicated, standard techniques as described by Seabrook, *et al.* (1989) (J. Seabrook, *et al.*, Molecular Cloning: A Laboratory Manual (2nd Ed.), Cold Spring Harbor Press (1989), hereinafter referred to as "Maniatis, *et al.* (1989)") and Ausubel, *et al.* (1987) (F.M. Ausubel, *et al.*, Current Protocols in Molecular Biology, New York: John Wiley & Sons (1987)), both of which are hereby incorporated in their entirety by this specific reference thereto, were utilized in the examples set out below. It should be understood that other methods of production of ssDNA, both by natural processes and by designed artificial methods using different enzyme products or systems, may also be utilized in connection with the method of the present invention and that the example set out herein are set out for purposes of exemplification as required by the Patent Statute and do not limit the intended scope of this disclosure.

The plasmid pcDNA3.1/Zeo+ was purchased from Invitrogen Corp. (Carlsbad, CA) and plasmid pBK-RSV from Statagene (La Jolla, CA). Oligodeoxynucleotides

(ODN) were synthesized by Midland Certified Reagent Co. (Midland, TX). Polymerase chain reactions (PCR) were carried out using Taq DNA polymerase purchased from Boehringer Mannheim Corp. (Indianapolis, IN) in a Robo-gradient thermal cycler (Stratagene (La Jolla, CA). Restriction endonucleases and T4 DNA ligase were obtained from Boehringer Mannheim Corp. (Indianapolis, IN). The ODNs used are listed in the attached Sequence Listing.

EXAMPLE 1. *In vivo* Synthesis of ss-cDNA in Eukaryotic Cells

The following *in vivo* experiments were designed to test whether the plasmids made in accordance with the present invention ssDNA in eukaryotic tissue culture cells.

Plasmid constructs. The ODNs were allowed to hybridize in 1 μl (5 μg/μl in water) in four separate tubes which were incubated at 70°C for 5 min and allowed to hybridize for 15 min at room temperature. Standard restriction endonuclease digests were carried out (EcoR I used as a negative control) with 10 units of enzyme in a total reaction volume of 15 μl and appropriate reaction buffers. DNA fragments were resolved in and isolated from agarose gels. The selection of positive clones on ampicillin plates was performed after transformation into competent XL1-Blue MRF' cells (Stratagene) as described by Maniatis, *et al.* (1989) and the accompanying instruction. After positive clones were picked, plasmid DNA was isolated using a commercially available plasmid isolation kit (Quiagen, Inc., Sant Caita, CA). Confirmation of DNA ligation was carried out by DNA sequencing.

The first plasmid was derived from pcDNA3.1/Zeo+ (Invitrogen Corp.) by digesting with restriction endonucleases Nhe I and Apa I, which are located in the multiple cloning site (MCS). The double-stranded oligodeoxynucleotide having compatible *Nhe* I and *Apa* I ends which is formed by annealing the synthetic, single stranded oligodeoxynucleotides Seq. ID 4/ODN-PMMV(+) and Seq. ID 5/ODN-PMMV(-) was ligated into the digested pcDNA3.1/Zeo+ to give pPMMV. This insert contains the Moloney Mouse leukemia virus reverse transcriptase (MoMuLV-RT) promoter region/primer binding site (PBS sequence). It also contains two *Not* I sites, unique to pMMV, and an *Mbo II* site. In this plasmid, and the plasmids derived from this construct, the strands designated (+) are positioned to be transcribed into RNA from the cytomegalovirus promoter of pcDNA3.1/Zeo+. The B plasmid, pssDNA-Express-B,

containing sites for convenient insertion of a sequence of interest to be expressed *in vivo* as ssDNA, was obtained by ligating the double-stranded sequence formed by annealing ODN-XB(+) and ODN-XB(-), which has *Not* I compatible overhangs, between the *Not* I sites of pMMV.

The structure of pss-DNA-Express-B is shown in Fig. 2. As set out above, this plasmid is constructed from the eukaryotic expression vector pcDNA3.1Zeo+. Mammalian cells containing the plasmid are selectable with the antibiotic zeocin. The position and general arrangement of key regions of the insert are shown in Fig. 2A and the specific arrangement of the sequences of the insert are shown in Fig. 2B. The exact positions of the structural features are shown in Fig. 2C. Transcription of the insert region is driven by the cytomegalovirus promoter and terminated in the BGH polyA region. The RNA transcript contains the MoMuLV core promoter along with some flanking regions of this promoter, the positions of which are indicated in Fig. 2. Reverse transcriptase synthesizes a copy of the (+) (top) strand, using the tRNA pro as a primer, beginning at the position of the core promoter. Digestion of the RNA strand by RNAse H releases a single-stranded DNA sequence containing the complementary inverted repeats IR-L and IR-R. Duplex formation by these repeats, shown in Fig. 9, creates a stem-loop with the sequence of interest in the loop. The stem contains a recognition site for Mbo II, GAAGA, positioned so that the enzyme, which cleaves 8/7 bases 3'to the GAAGA recognition site, releases the sequence of interest from the flanking vector sequences.

To obtain the plasmid pTest, from which the a sequence of interest that served as a control sequence is expressed, the double-stranded oligodeoxynucleotide formed by the annealed synthetic, single stranded oligodeoxynucleotides Seq. ID 11/ODN-Test(+) and Seq. ID 12/ODN-Test(-), which have compatible *Not* I ends was inserted between the *Not* I sites of pMMV. The plasmid expressing the telomeric repeat sequence pTelo as the sequence of interest, was obtained in similar fashion, inserting the annealed, *Not* I compatible oligodeoxynucleotides Seq. ID 13/ODN-Telo(+) and Seq. ID 14/ODN-Telo(-) between the Not I sites of pMMV. The latter linker sequence contains nine repeats of the vertebrate telomere sequence 5'-AGGGTT-3' (E.H. Blackburn, 350 Nature 569-573 (1991)). A third plasmid, termed pMN-new-link, was also constructed to give a more

stable stem-loop structure of 29 base pairs rather than the 27 base pair stem site of pMNV. pNM-new-link was formed by ligating two self-complementary ODN's, ODN-NM-new-link(+) and ODN-NM-new-link(-), between the two *Not* I sites within pMNV.

The second plasmid, pss-DNA-Express-A (Fig. 1), which contains the MoMuLV-RT (T.M. Shinnick, *et al.*, 293 Nature 543-548 (1981)) and restriction endonuclease genes, was derived from pBK-RSV (Stratagene), also using XL-1 Blue MRF' as the host strain. A mouse cell line expressing Moloney murine leukemia virus was obtained from the American Type Culture Collection (#CRL-1858). The virus RNA was isolated and prepared for reverse transcriptase-PCR (RT-PCR). A 2.4 kb fragment containing the coding sequence of MoMuLV-RT was PCR-amplified using primers as set out in Seq. ID 1/ODN-RT(-) (primer position at nucleotide #2545) and Seq. ID 2/ODN-RT(+) (primer position at nucleotide #4908) to produce a DNA fragment with a 5′-Sac I and a 3′-Hind III compatible end. The 2.4 kb product obtained includes the sequence of the MoMuLV genome between positions 2546 and 4908. The mature virus reverse transcriptase peptide is encoded by the sequence between positions 2337 and 4349 (Petropoulos, C.J., Retroviral taxonomy, protein structure, sequences and genetic maps, *in* J.M Coffin (Ed.), Retroviruses, 757, Appendix 2, New York: Cold Spring Harbor Press (1997)), but peptides truncated at the amino terminus retain full activity (N. Tanese, *et al.*, 85 Proc. Natl. Acad. Sci. USA 1777-1781 (1998)). The peptide encoded by this construct includes a part of the integrase gene, which follows the reverse transcriptase in the MoMuLV polyprotein, but is not relevant here such that the length of the construct was selected because of the availability of a convenient restriction site for cloning.

The bacterium *Moraxella bovis*, which encodes the restriction endonuclease MboII (H. Bocklage, *et al.*, 19 Nucleic Acids Res. 1007-1013 (1991)), was obtained from the American Type Culture Collection (ATCC#10900). Genomic DNA was isolated from *M. bovis* and the *Mbo* II gene was used as the template DNA in the PCR. A 1.2kb fragment containing the *Mbo* II gene was amplified by PCR using as primers Seq. ID 3/ODN-Mbo(+) (primer position at nucleotide #887) and Seq. ID 8/ODN-Mbo(-) (primer position at nucleotide #2206). These primers contain mismatches designed to introduce a a *Hind* III site into the 5′ primer and an *Xba* I site into the 3′ downstream primer. The 1.2kb DNA amplification product, copying the *M. bovis* genome between positions 888

and 2206, therefore contains the coding region for the Mbo II protein. The amplification product was digested with Hind III and Xba I.

The pBK-RSV vector was digested with Xba I and Nhe I, which removes the promoter region. The *Nhe* I end was converted to a *Sac* I end using the linker formed by annealed oligodeoxynucleotides Seq. ID 6/ODN-N>S(+) and Seq. ID 7/ODN-N>S(-). The reverse transcriptase and Mbo II amplimers were ligated through the Hind III sites and this construct was subsequently ligated between the *Sac* I and *Xba* I sites of pBK-RSV to produce pBK-RSV-RT/Mbo.

To insert a flexible linker between the reverse transcriptase and Mbo II domains of the polyprotein and to provide a tag useful for purification of the protein, the double-stranded sequence formed by annealing the oligodeoxynucleotides Seq. ID 9/ODN-HisPro(+) and Seq. ID 10/ODN-HisPro(-), encoding alternate histidine and proline amino acids. pBK-RSV-RT/Mbo was digested with Hind III, and the his-pro linker, with compatible *Hind* III ends, was inserted at the *Hind* III site to produce plasmid pBK-RSV-RT/Mbo-L and the orientation was confirmed by sequencing.

Sequencing pBK-RSV-RT/Mbo-L revealed a frame shift mutation at the 5'-end of the Mbo II domain. This mutation was corrected, and the extraneous part of the integrase gene of MoMuLV was removed simultaneously, by excising the fragment of the plasmid lying between the *Ase* I and *Bgl* II sites, which encodes the 5'-end of the *Mbo* II gene, the his-pro linker region, and the integrase gene fragment and replacing with an insert containing a modified his-pro linker and 5'-*Mbo* II gene fragment. The modified his-pro linker increased the number of histidines by one, to six, and included at the 5'-end a number of unique restriction sites. The 5'-end of the Mbo II gene was modified to replace the leucine at the N-terminus that was introduced by the mismatch in the PCR primer to the original methionine and to optimize codon usage for expression of this segment of the gene in mammalian cells. The repair construct was obtained by mutually-primed DNA synthesis from two templates, ODN-Rep(+) and ODN-Rep(-), that have complementary sequences of 16 bases at the 3'-ends. These oligodeoxynucleotides were annealed and extended with the modified SEQUENASE™ DNA polymerase enzyme (United States Biochemical Corp.). The double-stranded product was digested with Ase

I and Bgl II and inserted into the vector to give the plasmid pssDNA-Express-A (plasmid A).

The structure of pssDNA-Express-A is shown in Fig. 2. As set out above, to construct this plasmid, sequences encoding an active fragment of the MoMuLV reverse transcriptase and the *M. bovis* Mbo II restriction enzyme were cloned between the *Nhe* I and *Xma* I sites of the eukaryotic expression vector PBK-RSV. Transcription of the cloned region is driven by the RSV promoter and selection for transformed cells is carried out in the presence of the antibiotic G418 (neomycin). Reverse transcriptase and Mbo II are expressed as a single, bifunctional protein chain with the two functional domains separated by a short, histidine and proline rich linker.

Tissue culture studies. Stable and transient transfections were carried out by using lipofectant (Boehringer Mannhiem Corp.) using the manufacturer's accompanying instructions. All plasmid constructs were transfected into HeLa cell lines. Assays for ssDNA were performed by PCR and by dot-blot analyses 24-48 hours after transfection. Reverse transcriptase activity was assayed using the RT-PCR assay developed by Silver, et al. (Silver, J., *et al.* 21 Nucleic Acids Res. 3593-4 (1993)) after transfection with pssDNA-Express-A plasmid (Fig. 4, panel A). Individual colony isolates of stably substituted HeLa cell lines (A12 and B12) were additionally assayed for RT activity (Fig. 4B). The ss-cDNA was isolated from cells transfected 48-72-hr earlier. The ss-cDNA, which co-localizes with RNA, was carried out using trizol reagent (Gibco Life Technologies, Gaithersburg, MD). Assays for specific ss-cDNA species were carried out by both PCR based assays (Fig. 5 for pTest and Fig. 6 for pTelo) for internal fragment and by denatured single stranded gel electrophoresis with subsequent nylon blotting and probing with an internal biotin-labeled probe.

This experiment showed that human tissue culture cells (HeLa and Cos-7 cell lines), transfected with plasmids designed to synthesize a processed ss-cDNA, produced ss-cDNA of the predicted size. The cells were actually co-transfected with two plasmids, one plasmid carrying the RNA template encoding the stem-loop intermediate (Fig. 1A) and the other carrying the genes for reverse transcriptase and restriction endonuclease (Fig. 1C). Those skilled in the art, however, will recognize that a single plasmid including the RNA template for the stem-loop intermediate and the genes for reverse

transcriptase and restriction endonuclease can also be used for this purpose. The studies showed that the plasmid pc3.1/Zeo/N-M (Fig. 1A) synthesized an RNA encoded by the genetic element in Fig. 1B. The genetic element was regulated by the CMV promoter. The synthesized RNA supplied the template for the production of processed ss-cDNA.

The genes encoded in plasmid pBK-RSV-RT/MboL (Fig. 1C) included Moloney murine leukemia virus reverse transcriptase (which also has RNase H activity) and MboII restriction endonuclease. The synthesis of ss-cDNA in eukaryotic cell lines (HeLa) was confirmed using a "test" sequence contained within the linker portion of the plasmid construct pc3.1/Zeo/N-M. The test sequence provides a unique sequence for detection by hybridization using a complementary DNA probe. Using primers complementary to the 5′ and 3′ ends of the synthesized ss-cDNA and RNA isolated from the transfected cell lines (which is where one would expect to find the ss-cDNA fraction), PCR and agarose gel electrophoresis analyses showed that the ss-cDNA of predicted size was sensitive to S1 nuclease digestion. Appropriate negative and positive controls were included in the PCR analyses. PCR using the same 5′ and 3′ primers and the same isolated RNA amplified DNA of predicted size.

In addition, production of ss-cDNA was demonstrated by dot-blot analysis, in which a probe unique to the test sequence hybridized to the predicted isolates. It was deduced that the ss-cDNA formed a "stem-loop" intermediate *in vivo* because DNA of predicted size was produced if subsequent digestion by MboII had occurred. This latter result shows that the ss-cDNA formed a stem-loop intermediate and that the cloned MboII gene expressed a protein with enzymatic activity.

The vector pNM-New-Link (which after single stranded conversion contains the more stable 29 base pair stem structure) was used for in vitro experiments to demonstrate premature termination of the reverse transcriptase cDNA transcript at the 3′ aspect of the stem structure. As shown in Fig. 7, there was also some "read through" of this transcript through the stem structure as well (see the larger bands in Fig. 7). The sequence of interest produced from this premature termination is the second sequence of interest referred to herein

* * * * *

The experiments described above demonstrate two methods for producing ssDNA *in vivo* by multiple stepwise reactions using eukaryotic reverse transcriptase reactions and various cDNA priming reactions. These reactions were followed by formation of a "stem-loop" intermediate which can be used to eliminate any unwanted sequences either (a) upstream 5′ or downstream 3′ from a designed (and formed) "stem" after being subsequently cleaved by a restriction endonuclease or (b) by premature termination of the ss-cDNA from a 3′ sequence of interest.

Any nucleotide sequence of interest is produced by these methods in a eukaryotic cell. The sequence(s) of interest is cloned (or synthesized) between the designed inverted tandem repeats and represents the sequence in the "loop" after ssDNA production and subsequent stem-loop formation. In addition (or instead of), the sequence of interest is cloned (or synthesized) 3′ to the inverted repeats. The sequence(s) of interest to be produced are comprised of any base (i.e., A,T,G,C) composition as long as the sequence does not interfere with the formation of the stem of the stable stem-loop intermediate. Again, any restriction endonuclease may also be used to digest (or cleave) the stem portion of the stem-loop intermediate as long as the recognition site for that particular restriction endonuclease has been designed into the inverted tandem repeats.

Although described with reference to the figures and specific examples set out herein, those skilled in the art will recognize that certain changes can be made to the specific elements set out herein without changing the manner in which those elements function to achieve their intended respective results. For instance, the cassette described herein is described as being made up of a genetic element which comprises a sequence of interest and a tandem inverted repeat and a reverse transcriptase/RNAse H gene, each with appropriate promoters as described herein. Those skilled in the art will recognize that, for instance, the mouse Moloney leukemia virus reverse transcriptase gene described for use as the reverse transcriptase gene of the cassette can be replaced with other reverse transcriptase genes (the reverse transcriptase gene from Human immunodeficiency virus was one such gene which was noted above) and that promoters other than the CMV promoter may be used to advantage. Further, several restriction endonuclease genes are listed above, but those skilled in the art will recognize from this description that the list set out above is not exhaustive and that many other restriction endonuclease genes will function to advantage in connection with the present invention.

Similarly, the RSV promoter described as being used in connection with the restriction endonuclease genes set out herein is not the only promoter which may be used to advantage. All such changes and modifications which do not depart from the spirit of the present invention are intended to fall within the scope of the following non-limiting claims.

Table I: Oligodeoxynucleotides used in this work	
ODN-PMMV(+) 129 bases (#23)	5′-CTAGGTCGGCGGCCGCGAAGATTGGTGCGCACACACACAACGCGCA CCAATCTTCGCGGCCGCCGACCCGTCAGCGGGGGTCTTTCATTTGGGGG CTCGTCCGGGATCGGGAGACCCCTGCCCAGGGCC-3′
ODN-PMMV(-) 121 bases (#24)	5′-CTGGGCAGGGGTCTCCCGATCCCGGACGAGCCCCCAAATGAAAGAC CCCCGCTGACGGGTCGGCGGCCGCGAAGATTGGTGCGCGTTGTGTGTGT GCGCACCAATCTTCGCGGCCGCCGAC-3′
ODN-Test (+) 57 bases (#38)	5′-GGCCGGAAGATTGGGGCGCCAAAGAGTAACTCTCAAAGGCACGCGC CCCAATCTTCC-3′
ODN-Test (-) 57 bases (#39)	5′-GGCCGGAAGATTGGGGCGCGTGCCTTTGAGAGTTACTCTTTGGCGC CCCAATCTTCC-3′
ODN-Telo (+) 92 bases (#40)	5′-GGCCGGAAGATTGGGGCGTTAGGGTTAGGGTTAGGGTTAGGGTTAG GGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGCGCCCCAATCTTCC-3′
ODN-Telo (-) 92 bases (#41)	5′-GGCCGGAAGATTGGGGCGCCCTAACCCTAACCCTAACCCTAACCCT AACCCTAACCCTAACCCTAACCCTAACCCTAACGCCCCAATCTTCC-3′
ODN-XB(+) 51 bases	5′-GGCCTTGAAGAGCGGCCGCACTAACACCACCACAGTGCGGCCGCTC TTCAA-3′
ODN-XB(-) 51 bases	5′-GGCCTTGAAGAGCGGCCGCACTGTGGTGGTGTTAGTGCGGCCGCTC TTCAA-3′
ODN-RT (+) 32 bases (#13)	5′-GGGATCAGGAGCTCAGATCATGGGACCAATGG-3′
ODN-RT (-) 24 bases (#12)	5′-CTTGTGCACAAGCTTTGCAGGTCT-3′
ODN-N>S (+) 18 bases (#25)	5′-CTAGCGGCAAGCGTAGCT-3′
ODN-N>S (-) 10 bases (#26)	5′-ACGCTTGCCG-3′
ODN-Mbo (+) 30 bases (#16)	5′-CAATTAAGGAAAGCTTTGAAAAATTATGTC-3′
ODN-Mbo (-) 27 bases (#33)	5′-TAATGGCCCGGGCATAGTCGGGTAGGG-3′
ODN-HisPro (+) 43 bases (#36)	5′-AGCTGGATCCCCCGCTCCCCACCACCACCACCACCCTGCCCCT-3′
ODN-HisPro (-) 42 bases (#37)	5′-AGCAGGGGCAGGGTGGTGGTGGTGGTGGGGAGCGGGGGATCC-3′
ODN-Rep(+) 121 bases	5′-ATATCTATTAATTTTGGCAAATCATAGCGGTTATGCTGACTCAGGT GAATGCCGCGATAATTTTCAGATTGCAATCTTTCATCAATGAATTTCAG TGATGAATTGCCAAGATTGATGTTGC-3′
ODN-Rep(-) 111 bases	5′-GACGAGATCTCCTCCAGGAATTCTCGAGAATTCGGATCCCCCGCTC CCCACCACCACCACCACCACCCTGCCCCGCGGATGAAAAATTATGTGAG CAACATCAATCTTGGC-3′